

lncRNADB: a reference database for long noncoding RNAs

Paulo P. Amaral, Michael B. Clark, Dennis K. Gascoigne, Marcel E. Dinger* and John S. Mattick*

Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

Received August 16, 2010; Revised September 23, 2010; Accepted October 22, 2010

ABSTRACT

Large numbers of long RNAs with little or no protein-coding potential [long noncoding RNAs (lncRNAs)] are being identified in eukaryotes. In parallel, increasing data describing the expression profiles, molecular features and functions of individual lncRNAs in a variety of systems are accumulating. To enable the systematic compilation and updating of this information, we have developed a database (lncRNADB) containing a comprehensive list of lncRNAs that have been shown to have, or to be associated with, biological functions in eukaryotes, as well as messenger RNAs that have regulatory roles. Each entry contains referenced information about the RNA, including sequences, structural information, genomic context, expression, subcellular localization, conservation, functional evidence and other relevant information. lncRNADB can be searched by querying published RNA names and aliases, sequences, species and associated protein-coding genes, as well as terms contained in the annotations, such as the tissues in which the transcripts are expressed and associated diseases. In addition, lncRNADB is linked to the UCSC Genome Browser for visualization and Noncoding RNA Expression Database (NRED) for expression information from a variety of sources. lncRNADB provides a platform for the ongoing collation of the literature pertaining to lncRNAs and their association with other genomic elements. lncRNADB can be accessed at: <http://www.lncrnadb.org/>.

INTRODUCTION

The eukaryotic transcriptome is enormous, and comprises not only a large set of protein-coding messenger RNAs, but also large numbers of non-protein coding transcripts that have structural, regulatory or unknown functions (1). Noncoding RNAs include intergenic transcripts as well as a complex array of RNAs that overlap protein-coding loci on both strands (2–4), including promoter-associated RNAs (4), intronic RNAs (5), convergent and bi-directional transcripts (3), noncoding alternatively spliced isoforms of protein-coding genes (6) and mRNAs that also have regulatory roles as untranslated RNAs (7,8). While some of these RNAs are processed to form small RNA species such as microRNAs (miRNAs) and small nucleolar RNAs (snoRNAs), many can have diverse roles as primary or spliced long noncoding RNAs (lncRNAs) (9). In structural terms, lncRNAs range in size from approximately 100 to 100 000 bases, can be spliced or unspliced, polyadenylated or non-polyadenylated, nuclear or cytoplasmic, and are usually transcribed by RNA polymerase II and/or III (10).

The combination of different experimental methodologies to study the transcriptome of several species has resulted in a continuous discovery of novel transcripts (11), with the FANTOM project alone cataloguing more than 30 000 putative lncRNA transcripts in mouse tissues by full-length cDNA cloning (2). This provides a challenge for their molecular and functional characterization, as well as for their cataloguing. Although the subset of lncRNAs characterized to date corresponds to a small fraction of the long noncoding transcripts in multicellular organisms, growing interest in the field has resulted in a rapid increase in the characterization of individual lncRNAs in a variety of systems.

*To whom correspondence should be addressed. Tel: +61 7 3346 2079; Fax: +61 7 3346 2101; Email: m.dinger@imb.uq.edu.au
Correspondence may also be addressed to John S. Mattick. Tel: +61 7 3346 2079; Fax: +61 7 3346 2101; Email: j.mattick@uq.edu.au

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

© The Author(s) 2010. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

A decade ago less than a dozen lncRNAs were recognized and catalogued in all eukaryotes (12), but following the dramatic increase in their numbers, a recent catalogue listed approximately 40 functionally characterized lncRNAs in mammals alone (13). This list contained lncRNAs whose functions had been assessed either by loss- or gain-of-function experiments, and did not include many that have other types of evidence for functionality. For example, several studies have documented the cell type-specific and/or dynamic expression of hundreds of long ncRNAs in various developmental systems, such as embryonic stem cell differentiation (14) and oligodendrogenesis (15), finding in each case different subsets of differentially expressed lncRNAs.

These studies have indicated that lncRNAs comprise a class of bona fide gene products that have been largely unaccounted for in public databases such as RefSeq (16), UniGene (17) and the Mammalian Gene Collection (18). While there are databases cataloguing ncRNAs, these have been predominantly focused on well-validated classes of small RNAs and contain limited data pertaining to lncRNAs. Recently, a database with diverse information on imprinted RNAs was made available, which is largely comprised of small RNAs, such as piRNAs (piwi-interacting RNAs), but also contains imprinted lncRNAs, and is restricted to mammals (19). Our previous database of noncoding RNAs, RNADB (20), was also limited to mammals, and included all classes of regulatory RNAs, such as snoRNAs, piRNAs and miRNAs. Due to the rapid expansion in each of these classes, they are now independently curated in dedicated databases, including miRbase (21), piRNABank (22) and Sno/scaRNAbase (23). Consequently, there is a need for a dedicated database of lncRNAs that includes not only mammalian lncRNAs, but detailed annotations of lncRNAs from all eukaryotic species.

AIMS OF THE DATABASE

lncRNADB provides a central repository of known lncRNAs in eukaryotic cells (including those derived from viruses), their aliases and published characteristics. A well-collated library will greatly facilitate research of these poorly annotated genes. It also aims to decrease instances of replication and unknown identity by the inclusion of aliases such as *p15AS*, which was reported as a novel antisense RNA (24), but appears to be an unspliced isoform of a previously described ncRNA called *ANRIL* (25), and *Gomafu* (26), which is also known as *MIAT* (27) and *RNCR2* (28).

The centralization of lncRNADB enables integration with other resources including the UCSC Genome Browser (29) and the Noncoding RNA Expression Database (NRED) (30), which provide insight into genomic context and expression data. This ensures that researchers interested in lncRNAs can conveniently find a wide range of information regarding genes of interest from a single location.

Currently, the characteristics and functions of most lncRNAs are still unexplored, but it is expected that the

number of studied RNAs will rapidly increase by *in silico*, *in vitro* and *in vivo* characterization. By providing a simple interface accessible to public users, we aim to provide a tool for the scientific community that will allow existing entries to be updated, modified to improve accuracy, and new entries to be added by users. Following verification of new published data by the curators, the information will become available in lncRNADB.

DATABASE STRUCTURE

lncRNADB is available online at: <http://www.lncrnadb.org>. Users can search the database by lncRNA name, nucleotide sequence string, species, annotation status or through a full-text search with results being displayed for online perusal and available as a tab delimited file download.

Annotated entries include one or more literature references, annotations across a series of categories and a list of species in which the lncRNA is observed. Literature references and genomic coordinates are hyperlinked to PubMed and the UCSC Genome Browser, respectively. Other biological components related to the annotated lncRNAs, such as genomically-associated genes or interacting proteins, are also listed and briefly described in a separate table that provides links to the PubMed reference.

The lncRNADB database also links to NRED, our online expression analysis application for ncRNAs in mouse and human (30). This extends the database through access to relative expression levels of both the lncRNAs and their contextually related transcripts in various public microarray experiments, such as NCode data and the GNF atlas (31), as well as the Allen Brain Atlas (32), which includes mouse brain *in situ* expression data for over 800 expressed lncRNAs (33).

The application architecture consists of a Microsoft asp.NET 3.5 presentation layer (c#), c# 4.0 data model and application layer, and MySQL persistent storage.

QUERYING THE DATABASE

Querying the database is a matter of entering one or more search criteria (Figure 1A); a full or partial lncRNA name (or alias), a search string to interrogate each ncRNA's annotations or a species in which the lncRNA is known, or any combination of those. lncRNADB will return a list of matching entries whose full detail can be viewed by clicking on the lncRNA name.

To download the database values for the entire list as a file of tab delimited text, users can click on the 'Export Results' button under the search results.

lncRNADB CONTENT

Although lncRNAs have been defined previously as transcripts >200 nt (9), this was an arbitrary definition largely based on a convenient biochemical cutoff in RNA isolation protocols and the fact that it excluded most known small RNAs. In an effort to make a clear distinction to

small RNA species and to create a more biologically meaningful definition of lncRNAs, we consider lncRNAs as noncoding RNAs that may have a function as either primary or spliced transcripts, which is independent of processing into known classes of small RNAs, such as miRNAs, piRNAs and snoRNAs, while also excluding structural RNAs from classical housekeeping families (tRNAs, snoRNAs, spliceosomal RNAs, etc). Existing databases and archives, including Sno/scaRNAbase (23), miRBase (21), tRNAdb-CE (34) and piRNAbank (22), already represent such ncRNAs. However, some lncRNAs that are host genes for small RNA species (35,36), but may also have roles as regulatory lncRNAs, have been included in the annotations. An example is the *GAS5* lncRNA, which is a repressor of the glucocorticoid receptor but also encodes several intronic snoRNAs (37).

The database currently contains over 150 lncRNAs identified from the literature in around 60 different species. Each entry contains a comprehensive range of available information about the RNA, including sequences, structural information, genomic context, expression, subcellular localization, conservation, functional evidence and relevant ‘miscellaneous’ information (see e.g. in Figure 1B). As expected, most (~75%) of catalogued lncRNAs are from mammals, for which more transcriptomic data is available and which have been more intensively studied, but lncRNAs from vertebrates to single-celled eukaryotes have been included.

Among the entries in lncRNAdb, approximately 100 have functions directly tested by *in vitro* and/or *in vivo* experiments. For quick reference, these have been listed in Supplementary Table S1, highlighting the wide array of functional mechanisms and processes affected by

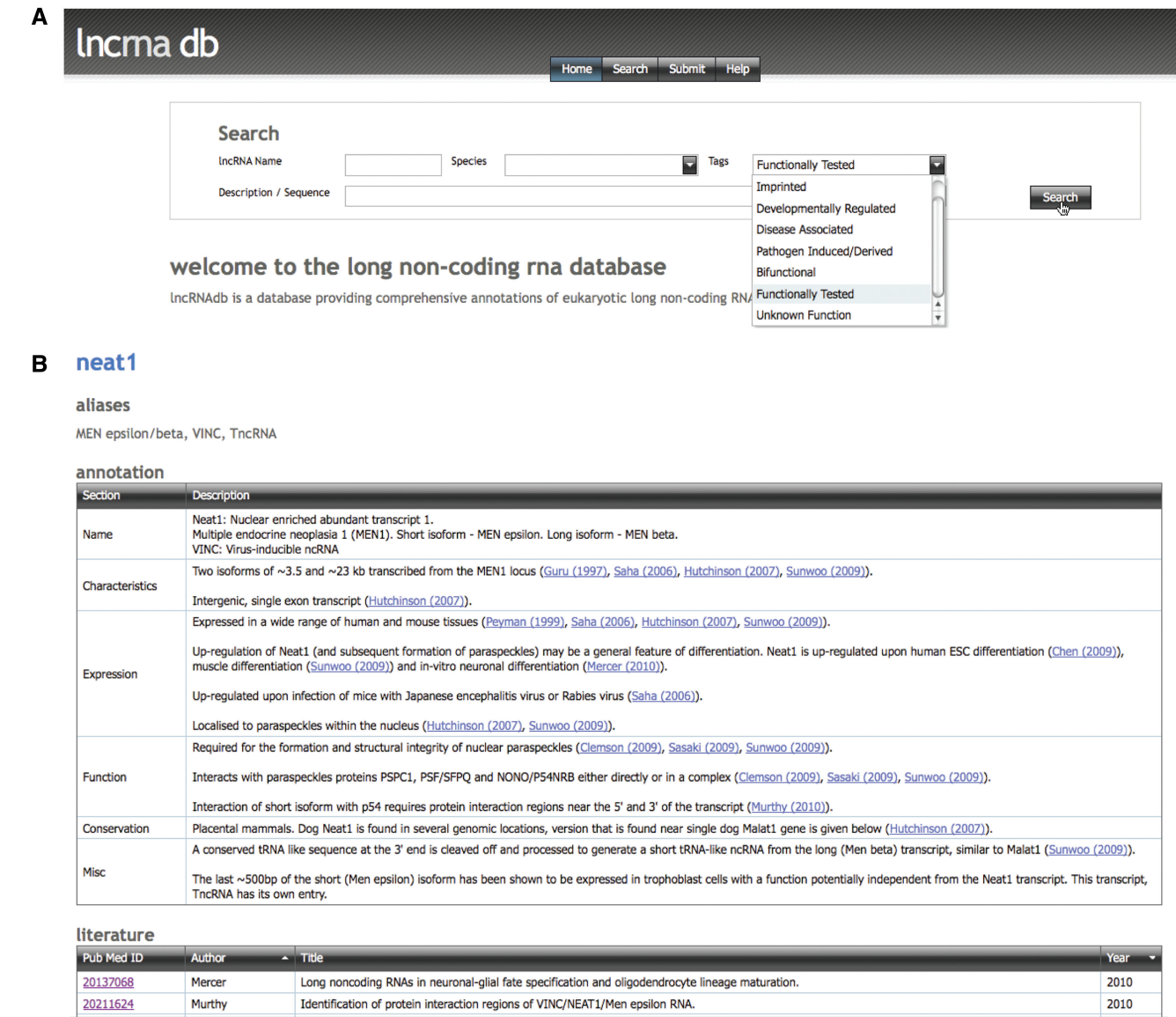


Figure 1. Representative screenshots of lncRNAdb showing (A) the search bar and (B) a lncRNA catalogued in lncRNAdb. Figure 1A shows the search fields available for querying the database and some of the pre-made descriptors available. Figure 1B depicts part of the annotation and references for the *Neat1* transcript.

lncRNAs. This list is not exhaustive, but it catalogues functional lncRNAs found not only in mammals but also lncRNAs tested in diverse eukaryotic species, such as *meiRNA* (38) in yeast, frq antisense RNAs in *Neurospora crassa* (39), rncs-1 in *Caenorhabditis elegans* (40), *IPSI* (41) in plants, *Xlsirts* in frogs (42) and *bereft* in *Drosophila* (43). In addition to RNAs expressed in normal physiological states, the database contains information about RNAs expressed in disease, and even lncRNAs derived from viruses and expressed in eukaryotic cells during infection (see below). Examples of lncRNA categories that can be used to limit queries are listed below.

Imprinted lncRNAs

lncRNAs are prevalent in imprinted regions where they can function to control imprinting and the expression of other genes from the locus, such as *Air* (44) and *Kncq1ot1* (45). Some imprinted lncRNAs are host genes for small RNAs, such as *Bsr* (46), where the lncRNA host has no known function; or has an independent function, e.g. the putative tumour suppressor *Meg3/Gtl2* (47).

Disease associated lncRNAs

Underscoring their importance in cellular functions, a growing number of lncRNAs have been implicated in a variety of diseases, including cancer. These are described in the database and include putative or confirmed cancer-associated lncRNAs, such as *NDM29* (48), *HOTAIR*, which regulates metastatic progression (49), and *H19*, which has been described as both an oncogene (50) and a tumour suppressor (51). We also include lncRNAs that have been implicated in neurological functions and diseases, including *BACE1AS* (52) that shows increased expression in Alzheimer's disease, and the *Drosophila* *hsr- ω* gene, which is induced by a variety of stresses and which has been shown to greatly increase protein polyglutamine-induced toxicity and neurodegeneration (53).

In addition, the observation that a substantial fraction of the genotypic variation underlying complex phenotypic traits occurs in noncoding regions, many of which are transcribed into discrete lncRNAs, has led to the appreciation that lncRNAs may play a central role in the molecular etiology of complex diseases (13). We have catalogued RNAs associated with these loci in lncRNAdb, such as *ANRIL*, a well-characterized lncRNA located in the complex genetic susceptibility locus *INK4b/ARF/INK4a* implicated in coronary artery disease, type 2 diabetes, periodontitis and cancer (24,54,55). We opted to also include uncharacterized lncRNAs linked to disease-susceptibility loci, as cataloguing these transcripts may facilitate their recognition as candidates for functional studies in normal and pathological conditions. Examples include *AK023948*, which is located in a susceptibility locus to papillary thyroid tumour (56), and *LOC285194*, which is located in a copy number alteration and loss-of-heterozygosity region in osteosarcoma (57).

Pathogen-induced or derived lncRNAs

Some lncRNAs are produced and modulated by pathogens or host cells during infection. These have been mostly omitted from other noncoding RNA catalogues, but incorporated in lncRNAdb. These include eukaryotic parasite transcripts such as the *Pincil* ncRNA family in the fungal plant-pathogen *Phytophthora infestans*, which are specifically upregulated during infection (58), and mammalian lncRNAs that are induced during viral infection (59), such as *Neat1/VINC* in the mouse brain (60), or produced in infected cells by oncogenic viruses, such as human herpesviruses (61). Indeed, the accumulating examples of lncRNAs encoded by viruses and expressed in eukaryotic cells have been annotated in lncRNAdb, because they can regulate cell function and are relevant to disease etiology. For instance, $\beta 2.7$ is a ~ 2.7 kb ncRNA encoded in the herpesvirus HCMV genome that is rapidly accumulated upon infection and has a fundamental role in preventing metabolic dysfunction and apoptosis of the host cell (62).

Bifunctional RNAs

An emerging class of genes is those encoding bifunctional RNAs, which can have multiple independent roles, such as acting as a regulatory lncRNA or being translated into a protein. Examples include well-known lncRNAs such as the co-activator *SRA* transcript, isoforms of which also encode a protein (8,63), and known protein coding genes, such as *p53*, whose transcripts also act as regulatory RNAs (7). In some cases, splicing isoforms are known to encode a protein, whereas specific splicing isoforms encode regulatory lncRNAs, such as the *LXRB/LXRBSV* isoform pair (6).

lncRNAs of unknown function

lncRNAdb includes RNAs that are well described in the literature but whose functions have yet to be identified. These regions are transcribed into lncRNAs that have been characterized to some extent at the structural and/or expression level, including dozens of lncRNAs that show tissue specificity and dynamic expression during development (64), as well as cellular localization, suggestive of undiscovered functionality. Likewise, the database also includes transcripts such as *PHO5* antisense lncRNA in yeast (65) where it is not yet established whether the functional role is conveyed by the lncRNA or its transcription.

FUTURE DIRECTIONS

The data in lncRNAdb will be extended through manual curation by the authors and submissions to the site by other researchers on an ongoing basis. The lowering cost and improved depths of RNA-sequencing methodologies is already enabling transcriptomics studies for alternative model- and non-model species, and it is expected that the identification and characterization of lncRNAs will follow, which may substantially increase the representation of non-mammalian species in the database. New functionality will be included with integration of the publicly

available successor to NRED (NRED2) early in 2011, giving access to not only additional microarray expression data but also transcriptomic RNA-seq data for contextually related long noncoding and coding genes, tools for visualization of the lncRNA and cross-referencing of experimental expression profiles. Finally, the absence of a centralized database of lncRNAs is reflected by the ad hoc naming of newly identified members, which are often not self-consistent and potentially confusing. With the number of functional lncRNAs likely to be very large, it will be important to establish a standardized nomenclature to minimize confusion and allow this emerging field to be as accessible as possible to all biologists.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Australian Research Council/University of Queensland co-sponsored Federation Fellowship (FF0561986; to J.S.M.); National Health and Medical Research Council of Australia Career Development Award (CDA631542; to M.E.D.); Queensland Government Department of Employment, Economic Development and Innovation Smart Futures Fellowship (to M.E.D.); Australian Research Council Postgraduate Awards (to P.P.A., M.B.C. and D.G.A.). Funding for open access charge: The University of Queensland.

Conflict of interest statement. None declared.

REFERENCES

- Amaral,P.P., Dinger,M.E., Mercer,T.R. and Mattick,J.S. (2008) The eukaryotic genome as an RNA machine. *Science*, **319**, 1787–1789.
- Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Nakamura,M., Oyama,R., Ravasi,T., Lenhard,B., Wells,C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
- Katayama,S., Tomaru,Y., Kasukawa,T., Waki,K., Nakanishi,M., Nakamura,M., Nishida,H., Yap,C.C., Suzuki,M., Kawai,J. *et al.* (2005) Antisense transcription in the mammalian transcriptome. *Science*, **309**, 1564–1566.
- Kapranov,P., Cheng,J., Dike,S., Nix,D.A., Duttagupta,R., Willingham,A.T., Stadler,P.F., Hertel,J., Hackermuller,J., Hofacker,I.L. *et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, **316**, 1484–1488.
- Nakaya,H.I., Amaral,P.P., Louro,R., Lopes,A., Fachel,A.A., Moreira,Y.B., El-Jundi,T.A., da Silva,A.M., Reis,E.M. and Verjovski-Almeida,S. (2007) Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol.*, **8**, R43.
- Hashimoto,K., Ishida,E., Matsumoto,S., Shibusawa,N., Okada,S., Monden,T., Satoh,T., Yamada,M. and Mori,M. (2009) A liver X receptor (LXR)-beta alternative splicing variant (LXRBSV) acts as an RNA co-activator of LXR-beta. *Biochem. Biophys. Res. Commun.*, **390**, 1260–1265.
- Candeias,M.M., Malbert-Colas,L., Powell,D.J., Daskalogianni,C., Maslon,M.M., Naski,N., Bourougaa,K., Calvo,F. and Fahraeus,R. (2008) p53 mRNA controls p53 activity by managing Mdm2 functions. *Nat. Cell Biol.*, **10**, 1098–1105.
- Chooniedass-Kothari,S., Hamedani,M.K., Auge,C., Wang,X., Carascossa,S., Yan,Y., Cooper,C., Vincett,D., Myal,Y., Jalaguer,S. *et al.* (2010) The steroid receptor RNA activator protein is recruited to promoter regions and acts as a transcriptional repressor. *FEBS Lett.*, **584**, 2218–2224.
- Mercer,T.R., Dinger,M.E. and Mattick,J.S. (2009) Long noncoding RNAs: insights into function. *Nat. Rev. Genet.*, **10**, 155–159.
- Dinger,M.E., Amaral,P.P., Mercer,T.R. and Mattick,J.S. (2009) Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. *Brief. Funct. Genomic Proteomic.*, **8**, 407–423.
- Huttenhofer,A. and Vogel,J. (2006) Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res.*, **34**, 635–646.
- Erdmann,V.A., Szymanski,M., Hochberg,A., de Groot,N. and Barciszewski,J. (1999) Collection of mRNA-like non-coding RNAs. *Nucleic Acids Res.*, **27**, 192–195.
- Mattick,J.S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet.*, **5**, e1000459.
- Dinger,M.E., Amaral,P.P., Mercer,T.R., Pang,K.C., Bruce,S.J., Gardiner,B.B., Askarian-Amiri,M.E., Ru,K., Solda,G., Simons,C. *et al.* (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.*, **18**, 1433–1445.
- Mercer,T.R., Qureshi,I.A., Gokhan,S., Dinger,M.E., Li,G., Mattick,J.S. and Mehler,M.F. (2010) Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. *BMC Neurosci.*, **11**, 14.
- Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
- Sayers,E.W., Barrett,T., Benson,D.A., Bolton,E., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Federhen,S. *et al.* (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **38**, D5–D6.
- Gerhard,D.S., Wagner,L., Feingold,E.A., Shenmen,C.M., Grouse,L.H., Schuler,G., Klein,S.L., Old,S., Rasooly,R., Good,P. *et al.* (2004) The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.*, **14**, 2121–2127.
- Zhang,Y., Guan,D.G., Yang,J.H., Shao,P., Zhou,H. and Qu,L.H. (2010) ncRNAimprint: a comprehensive database of mammalian imprinted noncoding RNAs. *RNA*, **16**, 1889–1901.
- Pang,K.C., Stephen,S., Dinger,M.E., Engstrom,P.G., Lenhard,B. and Mattick,J.S. (2007) RNAdB 2.0—an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res.*, **35**, D178–D182.
- Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
- Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, **36**, D173–D177.
- Xie,J., Zhang,M., Zhou,T., Hua,X., Tang,L. and Wu,W. (2007) Sno/scaRNAbase: a curated database for small nucleolar RNAs and cajal body-specific RNAs. *Nucleic Acids Res.*, **35**, D183–D187.
- Yu,W., Gius,D., Onyango,P., Muldoon-Jacobs,K., Karp,J., Feinberg,A.P. and Cui,H. (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, **451**, 202–206.
- Pasmant,E., Laurendeau,I., Heron,D., Vidaud,M., Vidaud,D. and Bieche,I. (2007) Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res.*, **67**, 3963–3969.
- Sone,M., Hayashi,T., Tarui,H., Agata,K., Takeichi,M. and Nakagawa,S. (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J. Cell Sci.*, **120**, 2498–2506.
- Ishii,N., Ozaki,K., Sato,H., Mizuno,H., Saito,S., Takahashi,A., Miyamoto,Y., Ikegawa,S., Kamatani,N., Hori,M. *et al.* (2006) Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J. Hum. Genet.*, **51**, 1087–1099.

28. Blackshaw,S., Harpavat,S., Trimarchi,J., Cai,L., Huang,H., Kuo,W.P., Weber,G., Lee,K., Fraioli,R.E., Cho,S.H. *et al.* (2004) Genomic analysis of mouse retinal development. *PLoS Biol.*, **2**, E247.
29. Rhead,B., Karolchik,D., Kuhn,R.M., Hinrichs,A.S., Zweig,A.S., Fujita,P.A., Diekhans,M., Smith,K.E., Rosenbloom,K.R., Raney,B.J. *et al.* (2010) The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.*, **38**, D613–D619.
30. Dinger,M.E., Pang,K.C., Mercer,T.R., Crowe,M.L., Grimmond,S.M. and Mattick,J.S. (2009) NRED: a database of long noncoding RNA expression. *Nucleic Acids Res.*, **37**, D122–D126.
31. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
32. Lein,E.S., Hawrylycz,M.J., Ao,N., Ayres,M., Bensinger,A., Bernard,A., Boe,A.F., Boguski,M.S., Brockway,K.S., Byrnes,E.J. *et al.* (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, **445**, 168–176.
33. Mercer,T.R., Dinger,M.E., Sunkin,S.M., Mehler,M.F. and Mattick,J.S. (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA*, **105**, 716–721.
34. Abe,T., Ikemura,T., Ohara,Y., Uehara,H., Kinouchi,M., Kanaya,S., Yamada,Y., Muto,A. and Inokuchi,H. (2009) tRNADB-CE: tRNA gene database curated manually by experts. *Nucleic Acids Res.*, **37**, D163–D168.
35. Tycowski,K.T., Shu,M.D. and Steitz,J.A. (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, **379**, 464–466.
36. Bortolin,M.L. and Kiss,T. (1998) Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. *RNA*, **4**, 445–454.
37. Kino,T., Hurt,D.E., Ichijo,T., Nader,N. and Chrousos,G.P. (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.*, **3**, ra8.
38. Watanabe,Y. and Yamamoto,M. (1994) *S. pombe* mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell*, **78**, 487–498.
39. Kramer,C., Loros,J.J., Dunlap,J.C. and Crosthwaite,S.K. (2003) Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. *Nature*, **421**, 948–952.
40. Hellwig,S. and Bass,B.L. (2008) A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes. *Proc. Natl Acad. Sci. USA*, **105**, 12897–12902.
41. Franco-Zorrilla,J.M., Valli,A., Todesco,M., Mateos,I., Puga,M.I., Rubio-Somoza,I., Leyva,A., Weigel,D., Garcia,J.A. and Paz-Ares,J. (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.*, **39**, 1033–1037.
42. Kloc,M., Wilk,K., Vargas,D., Shirato,Y., Bilinski,S. and Etkin,L.D. (2005) Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of *Xenopus* oocytes. *Development*, **132**, 3445–3457.
43. Hardiman,K.E., Brewster,R., Khan,S.M., Deo,M. and Bodmer,R. (2002) The bereft gene, a potential target of the neural selector gene cut, contributes to bristle morphogenesis. *Genetics*, **161**, 231–247.
44. Sleutels,F., Zwart,R. and Barlow,D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, **415**, 810–813.
45. Pandey,R.R., Mondal,T., Mohammad,F., Enroth,S., Redrup,L., Komorowski,J., Nagano,T., Mancini-Dinardo,D. and Kanduri,C. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell*, **32**, 232–246.
46. Royo,H., Basyuk,E., Marty,V., Marques,M., Bertrand,E. and Cavaille,J. (2007) Bsr, a nuclear-retained RNA with monoallelic expression. *Mol. Biol. Cell*, **18**, 2817–2827.
47. Zhang,X., Gejman,R., Mahta,A., Zhong,Y., Rice,K.A., Zhou,Y., Cheunsuchon,P., Louis,D.N. and Klibanski,A. (2010) Maternally expressed gene 3, an imprinted noncoding RNA gene, is associated with meningioma pathogenesis and progression. *Cancer Res.*, **70**, 2350–2358.
48. Castelnuevo,M., Massone,S., Tasso,R., Fiorino,G., Gatti,M., Robello,M., Gatta,E., Berger,A., Strub,K., Florio,T. *et al.* (2010) An Alu-like RNA promotes cell differentiation and reduces malignancy of human neuroblastoma cells. *Faseb J.*, **24**, 4033–4046.
49. Gupta,R.A., Shah,N., Wang,K.C., Kim,J., Horlings,H.M., Wong,D.J., Tsai,M.C., Hung,T., Argani,P., Rinn,J.L. *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464**, 1071–1076.
50. Matouk,I.J., DeGroot,N., Mezan,S., Ayeshe,S., Abu-lail,R., Hochberg,A. and Galun,E. (2007) The H19 non-coding RNA is essential for human tumor growth. *PLoS ONE*, **2**, e845.
51. Yoshimizu,T., Miroglio,A., Ripoche,M.A., Gabory,A., Vernucci,M., Riccio,A., Colnot,S., Godard,C., Terris,B., Jammes,H. *et al.* (2008) The H19 locus acts in vivo as a tumor suppressor. *Proc. Natl Acad. Sci. USA*, **105**, 12417–12422.
52. Faghihi,M.A., Modarresi,F., Khalil,A.M., Wood,D.E., Sahagun,B.G., Morgan,T.E., Finch,C.E., St. Laurent,G. 3rd, Kenny,P.J. and Wahlestedt,C. (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat. Med.*, **14**, 723–730.
53. Sengupta,S. and Lakhtia,S.C. (2006) Altered expressions of the noncoding hsromega gene enhances poly-Q-induced neurotoxicity in *Drosophila*. *RNA Biol.*, **3**, 28–35.
54. Holdt,L.M., Beutner,F., Scholz,M., Gielen,S., Gabel,G., Bergert,H., Schuler,G., Thiery,J. and Teupser,D. (2010) ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler. Thromb. Vasc. Biol.*, **30**, 620–627.
55. Yap,K.L., Li,S., Munoz-Cabello,A.M., Raguz,S., Zeng,L., Mujtaba,S., Gil,J., Walsh,M.J. and Zhou,M.M. (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell*, **38**, 662–674.
56. He,H., Nagy,R., Liyanarachchi,S., Jiao,H., Li,W., Suster,S., Kere,J. and de la Chapelle,A. (2009) A susceptibility locus for papillary thyroid carcinoma on chromosome 8q24. *Cancer Res.*, **69**, 625–631.
57. Pasic,I., Shlien,A., Durbin,A.D., Stavropoulos,D.J., Baskin,B., Ray,P.N., Novokmet,A. and Malkin,D. (2010) Recurrent focal copy-number changes and loss of heterozygosity implicate two noncoding RNAs and one tumor suppressor gene at chromosome 3q13.31 in osteosarcoma. *Cancer Res.*, **70**, 160–171.
58. Avrova,A.O., Whisson,S.C., Pritchard,L., Venter,E., De Luca,S., Hein,I. and Birch,P.R. (2007) A novel non-protein-coding infection-specific gene family is clustered throughout the genome of *Phytophthora infestans*. *Microbiology*, **153**, 747–759.
59. Mrazek,J., Kreutmayer,S.B., Grasser,F.A., Polacek,N. and Huttenhofer,A. (2007) Subtractive hybridization identifies novel differentially expressed ncRNA species in EBV-infected human B cells. *Nucleic Acids Res.*, **35**, e73.
60. Saha,S., Murthy,S. and Rangarajan,P.N. (2006) Identification and characterization of a virus-inducible non-coding RNA in mouse brain. *J. Gen. Virol.*, **87**, 1991–1995.
61. Swaminathan,S. (2008) Noncoding RNAs produced by oncogenic human herpesviruses. *J. Cell Physiol.*, **216**, 321–326.
62. Reeves,M.B., Davies,A.A., McSharry,B.P., Wilkinson,G.W. and Sinclair,J.H. (2007) Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science*, **316**, 1345–1348.
63. Foulds,C.E., Tsimelzon,A., Long,W., Le,A., Tsai,S.Y., Tsai,M.J. and O'Malley,B.W. (2010) Research resource: expression profiling reveals unexpected targets and functions of the human steroid receptor RNA activator (SRA) gene. *Mol. Endocrinol.*, **24**, 1090–1105.
64. Amaral,P.P. and Mattick,J.S. (2008) Noncoding RNA in development. *Mamm. Genome*, **19**, 454–492.
65. Uhler,J.P., Hertel,C. and Svejstrup,J.Q. (2007) A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc. Natl Acad. Sci. USA*, **104**, 8011–8016.